Could a Large Language Model be Conscious?

Abstract

There has recently been widespread discussion of whether large language models might be sentient. Should we take this idea seriously? I will break down the strongest reasons for and against. Given mainstream assumptions in the science of consciousness, there are significant obstacles to consciousness in current models: for example, their lack of recurrent processing, a global workspace, and unified agency. At the same time, it is quite possible that these obstacles will be overcome in the next decade or so. I conclude that while it is somewhat unlikely that current large language models are conscious, we should take seriously the possibility that successors to large language models may be conscious in the not-too-distant future.